Ecological Indicators 10 (2010) 264–273

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Quantifying human disturbance in watersheds: Variable selection and performance of a GIS-based disturbance index for predicting the biological condition of perennial streams

James A. Falcone *, Daren M. Carlisle, Lisa C. Weber

US Geological Survey, National Water-Quality Assessment Program, 413 National Center, 12201 Sunrise Valley Dr., Reston, VA 20192, United States

A R T I C L E I N F O

Article history:
Received 23 January 2009
Received in revised form 20 May 2009
Accepted 25 May 2009

Keywords:
Disturbance index
Watershed
GIS
Anthropogenic stress
Index performance

A B S T R A C T

Characterizing the relative severity of human disturbance in watersheds is often part of stream assessments and is frequently done with the aid of Geographic Information System (GIS)-derived data. However, the choice of variables and how they are used to quantify disturbance are often subjective. In this study, we developed a number of disturbance indices by testing sets of variables, scoring methods, and weightings of 33 potential disturbance factors derived from readily available GIS data. The indices were calibrated using 770 watersheds located in the western United States for which the severity of disturbance had previously been classified from detailed local data by the United States Environmental Protection Agency (USEPA) Environmental Monitoring and Assessment Program (EMAP). The indices were calibrated by determining which variable or variable combinations and aggregation method best differentiated between least- and most-disturbed sites. Indices composed of several variables performed better than any individual variable, and best results came from a threshold method of scoring using six uncorrelated variables: housing unit density, road density, pesticide application, dam storage, land cover along a mainstem buffer, and distance to nearest canal/pipeline. The final index was validated with 192 withheld watersheds and correctly classified about two-thirds (68%) of least- and most-disturbed sites. These results provide information about the potential for using a disturbance index as a screening tool for *a priori* ranking of watersheds at a regional/national scale, and which landscape variables and methods of combination may be most helpful in doing so.

Published by Elsevier Ltd.

1. Introduction

Stream ecosystems are profoundly influenced by human activities. Disturbances include point-source pollution, conversion of natural vegetation to developed land, nutrient and pesticide input from agricultural and urban sources, mining and mineral extraction operations, channel modification, and water impoundment. These activities often change the timing or amount of streamflow, increase runoff, erosion, and sedimentation, alter water temperature and chemistry, and introduce contaminants (Allan, 2004). The effect is a cumulative and often synergistic impact on water quality and quantity, habitat, and biotic assemblages (Stein et al., 2002). Identifying the nature and extent of human disturbances is a critical component in many ecological assessments.

Several studies have attempted to quantify the extent of human stressors and disturbances in an area with a single variable. The primary method of doing so is the creation of a "disturbance index" based on Geographic Information System (GIS)-derived data (Stein et al., 2002; Danz et al., 2005; Host et al., 2005; Wang et al., 2008). This allows for an objective integration of anthropogenic disturbance factors over large areas in a way that would be difficult to achieve with field studies. For ecological assessment, a disturbance index may be useful as a ranking or screening tool (Sanderson et al., 2002; Stein et al., 2002; Brown and Vivas, 2005; Wilhelm et al., 2005; Danz et al., 2007) or for an *a priori* classification of sites (Host et al., 2005; Wang et al., 2008). The accuracy of the index and success of classification are dependent on the availability, quality, time period, and scale of the GIS data, and the specific techniques employed.

Although a variety of approaches have been used in creating disturbance indices, several basic decisions are required: (1) which variable or combinations of variables should be included in the index, (2) how or if they should be weighted, and (3) how data values should be translated to an index score. It is clear that the performance of an index may vary significantly with the method chosen to score values (Blocksum, 2003). Given a set of GIS data it is possible to create any number of disturbance indices – based on

* Corresponding author. Tel.: +1 703 648 5008; fax: +1 703 648 6693.
E-mail address: jfalcone@usgs.gov (J.A. Falcone).

1470-160X/$ – see front matter. Published by Elsevier Ltd.
doi:10.1016/j.ecolind.2009.05.005

different variables, weightings, or methods of scoring or combination – which rank sites very differently. Further it is not necessarily known whether a multimetric index (an index composed of multiple variables) provides a better ranking than any individual variable, as in some cases a single variable (e.g. percent impervious surface) may adequately integrate multiple sources of human influence at some scales (Karr and Chu, 1999). Although biological data have been used in some cases to calibrate disturbance indices (Wang et al., 2008), most studies do not have access to externally classified watersheds for calibration.

Our goal was to compare and contrast methods of creating a disturbance index to determine those which provide the best discrimination between least- and most-disturbed watersheds that had previously been classified as such by the U.S. Environmental Protection Agency (USEPA) (2005). In addition we also evaluated the influence of various methods on the performance of the indices: (1) the use of correlated versus uncorrelated variables, (2) index scoring and weighting methods, (3) assessing which variables most strongly control the index, and (4) determining whether a multimetric index is a more powerful tool than any individual variable.

2. Methods

2.1. Study area and watershed classifications

We used data from 962 watersheds throughout the Western United States. These data were from watershed delineations for streams that had been randomly selected by the USEPA Environmental Monitoring and Assessment Program (EMAP; U.S. Environmental Protection Agency, 2005). Streams and attendant watersheds were therefore representative of 1:100,000 scale perennial streams throughout the region, which spans 12 states (Fig. 1). Watersheds ranged in size from 0.6 to 35,110 km^2 (median 44 km^2).

Sites were classified into three levels of overall disturbance by the EMAP. Sites classified as "reference" (hereafter REF) were considered to be representative of near natural or least disturbed conditions in their respective ecoregions. Conversely, sites classified as "most disturbed" (hereafter DIS) were considered to be representative of the most altered or modified by human activities. A third class, "intermediate" (hereafter MED), represented watersheds that have been altered to some intermediate degree. Sites were classified using an ecoregion-specific screening process based primarily on chemical and physical data collected on site (U.S. Environmental Protection Agency, 2005; Whittier et al., 2007). The process also included data on site-specific alterations based on aerial photography, field notes, and roundtable discussions among several principle USEPA investigators (Whittier et al., 2007). The final USEPA classification was based on independent classification by three investigators, one of whom also based his classification partly on GIS-derived data (percent agriculture and urban land cover in watershed and road density) (written communication, A. Herlihy, January 29, 2008). Because of the thoroughness of their review, and that only a portion of one-third of the decision-making process for classifying the sites was based on GIS data, we considered the USEPA evaluation to be as complete and independent an evaluation of the overall degree of physical and chemical disturbance in the watersheds for these sites as was possible to obtain. Consequently, our objective was to evaluate how well variations of disturbance indices based only on watershed-scale GIS data could reproduce this classification.

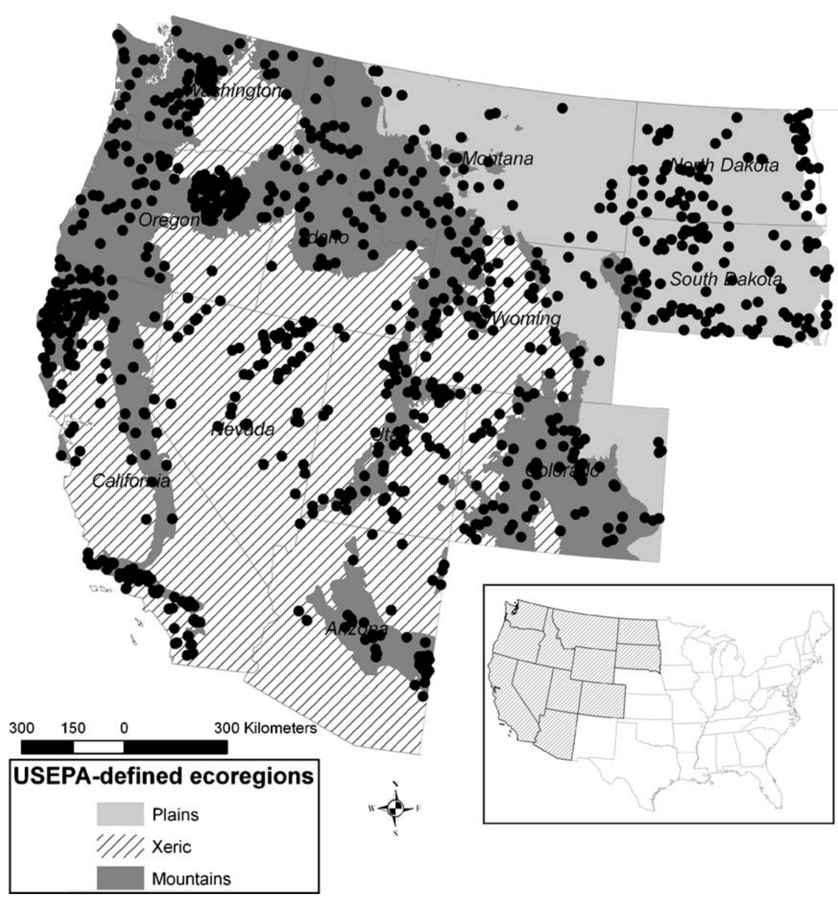

Fig. 1. U.S. Environmental Protection Agency (USEPA) sampling site locations of watersheds used in this study.

Table 1
Distribution of calibration and validation sites by ecoregion and type. Watershed characteristics are given as: median (minimum–maximum).

	All records	Calibration	Validation
	n = 962	n = 770	n = 192
Number by region			
Mountains	582 (60%)	454 (59%)	128 (67%)
Plains	194 (20%)	162 (21%)	32 (17%)
Xeric	186 (19%)	154 (20%)	32 (17%)
Number by type			
REF	218 (23%)	173 (22%)	45 (23%)
MED	554 (58%)	440 (57%)	114 (59%)
DIS	190 (20%)	157 (20%)	33 (17%)
Drainage area (sq km)	43.8 (0.6–35,110)	46.1 (0.6–35,110)	38.6 (0.6–14,220)
Road density (km/sq km)	0.79 (0.0–9.46)	0.79 (0.0–9.46)	0.83 (0.0–4.31)

We randomly selected 770 (80%) of the 962 watersheds for model calibration and reserved 192 (20%) to independently validate the accuracy of the final best index method. Watershed characteristics, as well as distribution by broad ecoregion and type (REF, MED, or DIS) were similar for calibration and validation groups (Table 1). Ecoregions used in this study were identical to those employed in the USEPA study (U.S. Environmental Protection Agency, 2005; Whittier et al., 2007): the western forested mountains (Mountains), xeric lands (Xeric), and the northern Great Plains (Plains).

2.2. Assembly of GIS data

We assembled GIS data from a number of sources (Table 2), all of which were available from national coverages. These represented measures of human population, infrastructure, land cover, pesticide and nutrient application, water impoundment, flow diversion, and point-source pollution. All data values were normalized to percent or per unit-area, and in the case of "distance" variables rank-transformed so that all variables ranged from low values (no or little disturbance) to high values (more disturbance). In all there were 33 disturbance variables, some of which measured to some degree the same broad landscape phenomenon (e.g. population density and housing unit density both generally measure "urbanization"), and several others were composites of other variables.

2.3. Evaluating the strength of individual variables

An assessment of the strength of individual variables in discriminating among watershed disturbance classes was used as part of the method for index development (described below). We used the non-parametric Kruskal–Wallis Chi-square (χ^2) test (Walpole and Myers, 1978), because most of the data distributions for the 33 variables tested here were not normally distributed. The Kruskal–Wallis χ^2 test is based on comparing the rank sums of two or more groups; large χ^2 values represent greater dissimilarity in rankings among groups. The test was run on the calibration dataset to assess the distinctness of the endpoint disturbance classes (REF and DIS) for each of the disturbance variables (Table 3).

2.4. Index development methods

Our goal was to identify the composition (i.e. set of disturbance variables) and configuration (e.g. scoring/weighting procedures) of an index that best discriminates between least- and most-disturbed watersheds, as previously classified by USEPA. Since testing every combination of variable, scoring, and weighting was not feasible, we limited *a priori* the number of combinations by evaluating four sets of variables (the set of all variables plus three subsets), for which we varied three scoring methods and three weightings. This would result in 36 indices to be tested (4 sets of variables $\times$ 3 scoring methods $\times$ 3 weightings). Having identified the best performing index with the calibration dataset, we evaluated the accuracy of the best set of variables, scoring

Table 2
Disturbance variables description and source.

Source of variable (reference; scale)	Variable abbreviation
Census 2000 (GeoLytics, 2001; Census block)	
Population density, persons/km^2	POPDEN
Housing unit density, housing units/km^2	HUDEN
Census 2000 TIGER roads (GeoLytics, 2001; 1:100,000)	
Road density in watershed, km/km^2	ROADDEN
National Land Cover Data 1992, enhanced (NLCDe) (Price et al., 2006; 30-m pixel)	
Urban land cover in watershed, percent	URB_BAS
Crops land cover in watershed, percent	CROPS_BAS
Pasture land cover in watershed, percent	PAST_BAS
Mining/transitional land cover in watershed, percent	MINING_BAS
Urban + crops land cover in watershed, percent	URBC_BAS
Urban + crops + pasture land cover in watershed, percent	URBCP_BAS
Urban land cover in mainstem 600-m buffer, percent	URB_MAINS
Urban + crops land cover in 600-m mainstem buffer, percent	URBC_MAINS
Urban + crops + pasture land cover in 600-m mainstem buffer, percent	URBCP_MAINS
USGS 1997 agricultural pesticide data (USGS, 2007; 1-km pixel)	
Sum of 43 most major pesticide compounds, kg/km^2	PESTIC
USGS (2002) nutrient data (Ruddy et al., 2006; 1-km pixel)	
Nitrogen from fertilizer and manure, kg/km^2	NITROGEN
Phosphorus from fertilizer and manure, kg/km^2	PHOSPH
Nitrogen + phosphorus, kg/km^2	N_P
National Resources Inventory-derived, 1992 (USGS, 2002; 1-km pixel)	
Land using NRI agricultural practice Surface Drainage Field Ditches, percent	NRI_DITCH
USGS NHDPlus 100 k streams (USEPA, 2006c; 1:100,000)	
Streams coded "Canal/Ditch/Pipeline" in watershed, percent	CANALS_BAS
Streams coded "Canal/Ditch/Pipeline" on mainstem, percent	CANALS_MAINS
Streams coded "Artificial Path" in watershed, percent	ARTPATH_BAS
Linear distance of sampling site to nearest canal/ditch/pipeline, m	DIST_CANAL_NEAR
Average linear distance of sampling site to all canals/ditches/pipelines, m	DIST_CANAL_AVG
EPA National Pollutant Discharge Elimination System (NPDES) data (USEPA, 2006b; 1:100,000)	
Density of discharge sites in watershed, sites/100 km^2	NPDES_ALL
Density of major discharge sites in watershed, sites/100 km^2	NPDES_MAJ
Linear distance of sampling site to nearest major NPDES site, m	DIST_NPDES_NEAR
Average linear distance of sampling site to all major NPDES site, m	DIST_NPDES_AVG
National Inventory of Dams (U.S. Army Corps of Engineers, 2006; 1:100,000)	
Dam storage in basin, liters $\times$ 1000/km^2	DAMSTOR
Dam density in watershed, dams/km^2	DAMDEN
Linear distance of sampling site to nearest dam, m	DIST_ANYDAM_NEAR
Average linear distance of sampling site to all dams, m	DIST_ANYDAM_AVG
Linear distance of sampling site to nearest major dam, m	DIST_MAJDAM_NEAR
Average linear distance of sampling site to all major dams, m	DIST_MAJDAM_AVG
USGS National Atlas, 2003 (U.S. Geological Survey, 2006; 1:2,000,000)	
Density of mineral operation sites, sites/100 km^2	MINOPS

methods, and weightings with the validation dataset.

Table 3
Kruskal–Wallis Chi-square (χ^2) values from the calibration dataset (n = 770) for testing the strength of separating least-disturbed and most-disturbed watershed classes. p-Values are all <0.0001 except where noted parenthetically. The original 33 variables were also grouped for Principal Components Analysis, creating the five synthetic variables (PCA class) that made up the Reduced-Synthetic dataset.

Disturbance variable	Chi-square value (p-value if $\geq$0.0001)	PCA class
HUDEN	66.9	POP
ROADDEN	62.1	POP
URBCP_MAINS	60.2	LC
URBCP_BAS	59.4	LC
POPDEN	56.3	POP
PESTIC	55.0	CHEM
PAST_BAS	53.4	LC
URBC_BAS	50.5	LC
URBC_MAINS	49.7	LC
URB_BAS	46.2	LC
URB_MAINS	44.4	LC
NITROGEN	41.5	CHEM
N_P	41.0	CHEM
CROPS_BAS	41.0	LC
PHOSPH	39.3	CHEM
DAMSTOR	29.0	FLOW
DIST_CANAL_NEAR	28.3	FLOW
CANALS_BAS	28.1	FLOW
DIST_ANYDAM_NEAR	27.4	FLOW
DIST_CANAL_AVG	26.2	FLOW
DIST_ANYDAM_AVG	24.7	FLOW
DAMDEN	24.6	FLOW
MINOPS	22.2	PS
DIST_MAJDAM_NEAR	21.3	FLOW
DIST_MAJDAM_AVG	21.2	FLOW
NPDES_ALL	19.8	PS
CANALS_MAINS	19.4	FLOW
NRI_DITCH	16.7	FLOW
DIST_NPDES_NEAR	14.38 (0.0002)	PS
DIST_NPDES_AVG	14.35 (0.0002)	PS
NPDES_MAJ	14.33 (0.0002)	PS
MINING_BAS	5.75 (0.0165)	LC
ARTPATH_BAS	1.86 (0.1725)	FLOW

method, and weighting on an independent set of data (validation set, n = 192). These steps are described in more detail below.

2.4.1. Subsetting variables: Reduced-Original dataset

A reduced dataset was developed (hereafter known as the Reduced-Original dataset) from the original set of 33 variables. We first eliminated variables that were correlated (Spearman rho > 0.7). For any two correlated variables, the variable with the highest χ^2 value (from Table 3) was selected. This resulted in a reduced set of 11 variables: HUDEN (housing unit density), ROADDEN (road density), PESTIC (pesticide application), URBCP_MAINS (percent urban + agriculture land cover in 600 m mainstem buffer), DIST_CANAL_NEAR (site distance to nearest canal), DAMSTOR (dam storage), MINOPS (mineral operations locations), NPDES_ALL (permitted discharge locations), NRI_DITCH (percent ditches), DIST_NPDES_NEAR (site distance to nearest major discharge location), MINING_BAS (percent mining land cover in watershed).

Variables where a large number of values are identical are not likely to provide information that helps differentiate sites (U.S. Environmental Protection Agency, 2006a), therefore variables that had values of 0 for more than 90% of sites were also eliminated (NRI_DITCH and MINOPS). Additionally, variables that were the least distinct between REF and DIS watersheds (χ^2 value < 25) were eliminated (MINING_BAS, NPDES_ALL, and DIST_NPDES_NEAR). This final selection resulted in a group of six variables: HUDEN, ROADDEN, PESTIC, URBCP_MAINS, DIST_CANAL_NEAR and DAMSTOR.

2.4.2. Subsetting variables: Reduced-Synthetic dataset

A second reduced dataset (hereafter known as Reduced-Synthetic dataset) was developed from synthetic variables based on Principal Components Analysis (PCA). We followed the basic method described by Danz et al. (2007). This consisted of first grouping the 33 variables into five categories based on professional judgment: population/roads (POP; 3 variables), land cover (LC; 9 variables), pesticides/nutrients (CHEM; 4 variables), flow impoundments/diversions (FLOW; 12 variables), and point-source pollution (PS; 5 variables). Table 3 shows which original variables were included in each category. For each category we then performed a PCA, which transforms the original data into a set of new variables which often summarize the majority of the variation of the input data in the first few Principal Components (PCs; Danz et al., 2007). We used the correlation matrix to perform the PCA instead of the covariance matrix because the variables were measured in various scales and units. We interpreted the PCs by examining the factor loadings against the original input variables. Components beyond the first PC were not easily interpretable, therefore the first PC was selected for use in all five cases as the best indicator of overall stress for that category. Values for the first PC were normalized to range between 0 and 1 for each category, creating five new disturbance variables corresponding to each category, given the following abbreviations: PCA1_POP, PCA2_LC, PCA3_CHEM, PCA4_FLOW, and PCA5_PS.

2.4.3. Subsetting variables: Redundant dataset

The third reduced dataset we created (hereafter known as Redundant dataset) mimicked the Reduced-Original dataset, but intentionally included additional variables that were correlated (similar measures of the same phenomenon) in order to test the effect of redundancy in an index. All variables that had Spearman rho > 0.7 when correlated with any of the six variables in the Reduced-Original dataset were included, as well as those original six. This resulted in a dataset with 22 variables.

2.4.4. Scoring methods

For each of the four datasets, we evaluated three methods of scoring data values. Translation of data values into a disturbance score has been done in a number of ways, often according to how the data are organized. For example, a global-scale grid-based approach might assign disturbance points based on the presence/absence of a disturbance (Sanderson et al., 2002). A common approach in other studies is assignment of disturbance points based on percentile or relative values for the population of watersheds being studied (Host et al., 2005; Wilhelm et al., 2005). We used three variations of this latter approach.

The first method was termed the range-standardize scoring method. In this method data values for a variable were standardized to range from 0 to 10 to produce raw disturbance scores. The raw scores were multiplied by a weighting factor (see below), then summed for all variables included in the index. The summed values were then standardized again so that the final index values ranged from 0 (lowest disturbance) to 10 (highest disturbance). This method has been used in a number of studies for multimetric index creation (McMahon and Cuffney, 2000; Falcone et al., 2007), and is a method in which every data value that is greater than 0 contributes to the overall disturbance score.

Our second scoring approach was termed the Percentile > 0 scoring method. This approach assigned raw disturbance scores based on percentile thresholds of non-zero values. If a data value was >0 a raw score of 1 was assigned for values >= 1st and <= 20th percentile, 2 for 21–40th percentile, 3 for 41–60th percentile, 4 for 61–80th percentile, and 5 for 81–100th percentile. Raw scores were multiplied by a weighting factor, summed, then rescaled to range from 0 to 10. This method is similar to the previous method in that all data values greater than 0 contribute to the overall disturbance score, but the effect of outlier values is minimized by binning the values into five groups.

Our third scoring method was termed the Percentile > 50 scoring method. This approach assigned disturbance scores only if data values were greater than the median value. A raw score of 1 was assigned for values >50 and <= 60th percentile, 2 for 61–70th percentile, 3 for 71–80th percentile, 4 for 81–90th percentile, and 5 for 91–100th percentile. Raw scores were multiplied by a weighting factor, summed, then rescaled to range from 0 to 10. With this method low disturbance values were assumed to not contribute to the overall disturbance; e.g. there may be roads in the watershed, but unless they cross the median road density threshold they are assumed to contribute no disturbance.

2.4.5. Weighting methods

Some variables in a disturbance index are likely to be "more important" than others with respect to the magnitude of stress they may cause to stream ecosystems, and should therefore be weighted accordingly. Several studies have weighted index variables based on either professional judgment (e.g. Sanderson et al., 2002; Stein et al., 2002) or on a statistically derived measure from the data themselves (e.g. multivariate ordination; Wang et al., 2008).

In this study we tested three methods. The first method was no weighting, i.e. all variables given equal weight of 1.0. The second method, termed the χ^2 weighting method, was based on weighting the χ^2 values given in Table 3, range-standardized from 0 to 1 (Table 4). With this weighting method the weights correspond to how well that variable individually discriminated between the least- and most-disturbed watersheds.

Table 4
Values used for the weighting schemes tested. The Chi-square weights are simply the values from Table 3 rescaled from 0 to 1. The Principal Components (PC) weights were derived from weighted loadings of a Principal Components Analysis of all variables, rescaled from 0 to 1.

Disturbance variable	Chi-square (χ^2) weight	PC weight
HUDEN	1.00	0.93
PCA1_POP	0.98	1.00
ROADDEN	0.93	0.47
URBCP_MAINS	0.90	0.90
URBCP_BAS	0.88	0.88
PCA2_LC	0.84	0.99
POPDEN	0.84	0.94
PESTIC	0.82	0.63
PAST_BAS	0.79	0.29
URBC_BAS	0.75	0.85
URBC_MAINS	0.73	0.93
URB_BAS	0.68	0.85
URB_MAINS	0.65	0.72
PCA3_CHEM	0.61	0.90
NITROGEN	0.61	0.89
N_P	0.60	0.89
CROPS_BAS	0.60	0.71
PHOSPH	0.58	0.87
PCA4_FLOW	0.48	0.60
DAMSTOR	0.42	0.11
PCA5_PS	0.41	0.75
DIST_CANAL_NEAR	0.41	0.44
CANALS_BAS	0.40	0.40
DIST_ANYDAM_NEAR	0.39	0.22
DIST_CANAL_AVG	0.37	0.43
DIST_ANYDAM_AVG	0.35	0.19
DAMDEN	0.35	0.00
MINOPS	0.31	0.07
DIST_MAJDAM_NEAR	0.30	0.32
DIST_MAJDAM_AVG	0.30	0.33
NPDES_ALL	0.28	0.16
CANALS_MAINS	0.27	0.06
NRI_DITCH	0.23	0.23
DIST_NPDES_NEAR	0.19	0.56
DIST_NPDES_AVG	0.19	0.57
NPDES_MAJ	0.19	0.52
MINING_BAS	0.06	0.00
ARTPATH_BAS	0.00	0.00

The third weighting method was termed the PC weighting method. In this method, the derived weights are based on variable importance as judged from principal components loadings and eigenvalues (Yang and Shahabi, 2004). PCA for all variables using the correlation matrix was performed and the first six PCs (proportion of variance = 0.75) were deemed to be interpretable (this PCA differed from the one used to create the Reduced-Synthetic dataset, in which variables were first grouped into five categories, then PCA run on each category). Component one loaded strongly to agriculture, two and three to urbanization, four to point-source pollution, five to diversions/canals, and six to dams. Loadings of each variable for each component were range-standardized (0–1), then multiplied by the eigenvalue for that component, giving loadings which were more heavily weighted for the first PCs. These values were then summed across all six components for each variable and range-standardized again to produce "importance scores" ranging from 0 to 1 (Table 4). In this way variables which had the highest loadings for the most important PCs were given highest weights.

Applying all combinations of the above factors created the 36 indices tested in this study (4 sets of variables × 3 scoring methods × 3 weightings). All indices were evaluated by comparing the REF and DIS watersheds with the Kruskal–Wallis χ^2 test (test equality of rankings). Larger (absolute) values for the χ^2 test indicated better ability to separate REF and DIS watersheds based on the USEPA classification. Using the χ^2 test also allowed comparison of index results to the χ^2 values for the individual variables (Table 3). The final best index method/variable combination that resulted was then applied to all records in the validation dataset ($n = 192$).

2.5. Testing against validation data

We evaluated the accuracy with which the selected index could be used to classify an independent set of watersheds into reference and disturbed categories. We calculated the disturbance score for each watershed in the validation dataset, then used the following procedure to classify them as least- or most-disturbed. Watersheds were separated by ecoregion, then ecoregion-specific thresholds were applied according to the 75th percentile rank of the reference sites from the calibration data (see Fig. 4). For example, if a validation site score for a Mountain-region site was higher than 2.3 it was classified as DIS, otherwise it was classified as REF. Those classifications were then compared against the USEPA classification for the known REF and DIS sites. This resulted in either a correct or incorrect classification for those sites, from which a percent error could be calculated.

3. Results

The primary method for evaluating indices in this study was the Kruskal–Wallis χ^2 statistic produced by comparing our disturbance index scores from the USEPA-classified REF and DIS sites. The statistic is an estimate of the degree to which an index differentiates between watersheds known to be most-disturbed and watersheds known to be least-disturbed. The χ^2 values for the indices based on reduced data (indices A–AC; Table 5) were almost uniformly higher than χ^2 values for indices using all of the original variables (indices AD–AM). Comparing values in Table 5 allows evaluation of how changing only one element of an index affected performance. For example, a comparison of indices A, K, U and AD shows that all else being equal (i.e. scoring and weighting methods), an index incorporating all 33 original variables had poorer performance than indices that had been reduced (at least some redundancy removed).

The significance of differences between Chi-square values with equal degrees of freedom may be evaluated using a simple

Table 5
Results of index testing. Higher Chi-square values and represent better ability to separate least- and most-disturbed watersheds (d.f. = 1; all p-values <0.0001). Differences >4 between Chi-square values are significantly different at the 0.05 level. The dataset and methods from index E (bolded) were also tested against the validation dataset (192 watersheds). Shaded rows represent the scoring method "Percentile >0", as discussed in the text.

Index designation	Variables in index	Scoring method	Weighting	Chi-square value	Median REF score	Median DIS score
A	Reduced-Original (6)	Range-standardize	Equal	84.4	0.16	0.83
B	Reduced-Original (6)	Range-standardize	χ^2	84.3	0.19	0.86
C	Reduced-Original (6)	Range-standardize	PC	85.0	0.27	0.97
D	Reduced-Original (6)	Percentile >0	Equal	88.2	1.33	4.67
E	**Reduced-Original (6)**	**Percentile >0**	$\boldsymbol{\chi^2}$	**90.3**	**1.25**	**5.42**
F	Reduced-Original (6)	Percentile >0	PC	88.9	1.84	5.70
G	Reduced-Original (6)	Percentile >50	Equal	70.2	0.00	2.07
H	Reduced-Original (6)	Percentile >50	χ^2	71.3	0.00	2.17
J	Reduced-Original (6)	Percentile >50	PC	73.1	0.00	2.70
K	Reduced-Synthetic (5)	Range-standardize	Equal	72.5	0.06	0.16
L	Reduced-Synthetic (5)	Range-standardize	χ^2	72.2	0.09	0.23
M	Reduced-Synthetic (5)	Range-standardize	PC	72.6	0.09	0.22
N	Reduced-Synthetic (5)	Percentile >0	Equal	85.2	2.50	5.83
P	Reduced-Synthetic (5)	Percentile >0	χ^2	88.2	2.88	6.11
Q	Reduced-Synthetic (5)	Percentile >0	PC	87.0	3.08	6.39
R	Reduced-Synthetic (5)	Percentile >50	Equal	79.1	0.40	3.20
S	Reduced-Synthetic (5)	Percentile >50	χ^2	77.0	0.51	3.37
T	Reduced-Synthetic (5)	Percentile >50	PC	78.3	0.52	3.51
U	Redundant (22)	Range-standardize	Equal	77.0	0.10	2.20
V	Redundant (22)	Range-standardize	χ^2	79.0	0.13	1.66
W	Redundant (22)	Range-standardize	PC	76.1	0.21	1.36
X	Redundant (22)	Percentile >0	Equal	73.5	1.36	4.56
Y	Redundant (22)	Percentile >0	χ^2	78.7	1.57	5.00
Z	Redundant (22)	Percentile >0	PC	74.7	2.40	5.58
AA	Redundant (22)	Percentile >50	Equal	63.5	0.32	2.04
AB	Redundant (22)	Percentile >50	χ^2	68.5	0.31	2.06
AC	Redundant (22)	Percentile >50	PC	64.3	0.66	2.69
AD	All (33)	Range-standardize	Equal	66.5	0.11	1.76
AE	All (33)	Range-standardize	χ^2	75.9	0.12	1.60
AF	All (33)	Range-standardize	PC	74.3	0.21	1.21
AG	All (33)	Percentile >0	Equal	71.2	1.27	4.15
AH	All (33)	Percentile >0	χ^2	78.1	1.38	4.83
AJ	All (33)	Percentile >0	PC	74.7	2.43	5.73
AK	All (33)	Percentile >50	Equal	54.2	0.50	1.82
AL	All (33)	Percentile >50	χ^2	66.8	0.26	1.91
AM	All (33)	Percentile >50	PC	63.9	0.62	2.68

transformation to z-scores (Knepp and Entwisle, 1969). This is given as

$$z = \frac{\chi_1^2 - \chi_2^2}{2(\sqrt{\nu})}, \quad \text{where}\ \nu = \text{degrees of freedom}. \qquad (1)$$

The z-score may then be evaluated against a standard normal distribution. In the context of the Chi-square values discussed in this paper, where d.f. = 1, this indicates that differences >4 between any two Chi-square values are significant at the 0.05 level, and differences >3.2 are significant at the 0.10 level.

Inclusion of correlated variables tended to decrease index performance. In every case indices based on the Reduced-Original dataset (indices A–J) performed better than their Redundant counterparts (indices U–AC). The Reduced-Original data also performed better in every case than indices using all original 33 variables; and in nearly every case the Reduced-Synthetic dataset (indices K–T) outperformed both the Redundant dataset and the original 33 variables as well. This strongly suggested that removing redundant or highly correlated variables from the index calculation was beneficial.

The scoring method appeared to have a considerable effect on index performance. For example, index E (χ^2 = 90.3) and index H (χ^2 = 71.3) are identical in computation except for the scoring method (Percentile >0 and Percentile >50, respectively). However, the boxplots of data distributions of REF and DIS scores for those two indices were substantially different, as noted by the small amount of overlap in the middle 50-percent of the data inside the box (Fig. 2). For each combination of variables and weighting, the Percentile >0 scoring method always outperformed the Percentile >50 scoring method (comparing index pairs D/G, E/H, F/J, N/R, P/S, Q/T, X/AA, Y/AB, Z/AC, AG/AK, AH/AL, and AJ/AM). In 9 of 12 comparisons the Percentile >0 scoring method also outperformed the range-standardize scoring

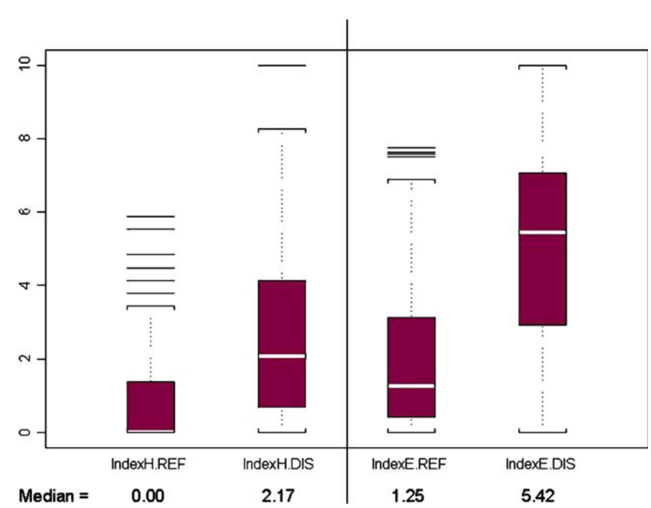

Fig. 2. Box plots of disturbance index scores (y-axis) from two indices (E and H), showing distribution of USEPA-classified least-disturbed (REF; n = 173) and most-disturbed (DIS; n = 157) sites for each. Indices were identical except for the scoring method. Index E (right side) shows good separation of REF and DIS sites, while index H (left side) shows much poorer separation.

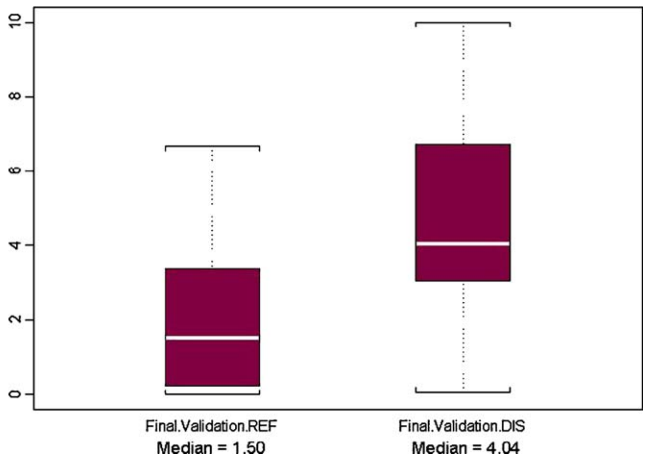

Fig. 3. Box plots of disturbance index scores (y-axis) from best index method, based on index E, applied to validation dataset. Left box plot are scores for sites USEPA classified as least-disturbed (REF; $n = 45$); right box plot are scores for sites classified as most-disturbed (DIS; $n = 33$).

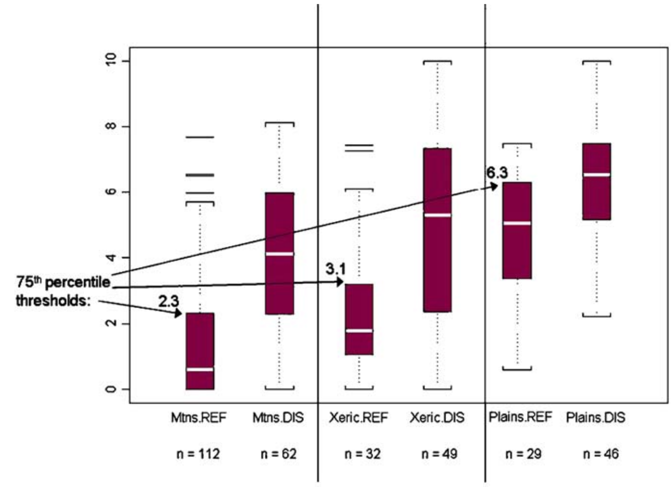

Fig. 4. Box plots of disturbance index scores (y-axis) by region for final calibration data, showing distribution of USEPA-classified least-disturbed (REF; $n = 173$) and most-disturbed (DIS; $n = 157$) sites for Mountains (left), Xeric (middle), and Plains (right) regions. 75th Percentile thresholds shown are values used for assigning classification for validation sites from final index model.

method (index pairs A/D, B/E, C/F, K/N, L/P, M/Q, AD/AG, AE/AH, and AF/AJ).

The weighting method also appeared to influence index performance, although generally less so than the scoring method. Indices that used χ^2 weighting generally outperformed indices that had equal weighting (9 of 12 comparisons: index pairs D/E, G/H, N/P, U/V, X/Y, AA/AB, AD/AE, AG/AH, and AK/AL). PC weighting likewise nearly always outperformed equal weighting (10 of 12 comparisons: index pairs A/C, D/F, G/J, K/M, N/Q, X/Z, AA/AC, AD/AF, AG/AJ, and AK/AM). For our best scoring method (Percentile > 0; shaded rows in Table 5), indices using either χ^2 or PC weighting were always better than indices with equal weighting.

The best combination of dataset, scoring method, and weighting (Reduced-Original dataset, Percentile > 0 scoring, and χ^2 weighting, respectively), was provided by index E, as judged by its χ^2 value. When applied to the validation watersheds, index E varied significantly ($F = 29.0$, $p < 0.0001$; Fig. 3) between REF and DIS. Using the thresholds derived from regional distribution of REF and DIS scores of the calibration data (Fig. 4), 71.1% of REF validation watersheds were correctly classified (32/45), 63.6% of DIS validation watersheds were correctly classified (21/33), and overall 67.9% of validation watersheds were correctly classified (53/78). The kappa coefficient, representing the agreement above what would be expected by chance, was 0.40.

4. Discussion

Few studies have had the opportunity to test results of disturbance index creation against rigorously derived independent validation data. Being able to do so allows the effect of specific aspects of index creation to be systematically tested and observed.

4.1. Effect of correlated variables

The collection of data which may be redundant or correlated is a common occurrence. Several solutions are possible: one is to pick the variable which is believed to be the "best" representative of each class (e.g. agriculture or water impoundment); another solution is to reduce the data into new classes using a data ordination technique (e.g. PCA); and another solution is to do nothing and use all variables without reduction. No studies to our knowledge have formally tested the effect of different solutions to data reduction on index creation.

In this study we tested and compared all three options. Removing correlated variables created an improved index in every case where it was explicitly tested (comparing Reduced-Original to Redundant dataset indices), and in nearly every case indices created from any reduced dataset were superior to using all original redundant variables (Fig. 5a). The implication is that because redundancy is likely to skew the index calculation in favor of certain classes and may create unnecessary noise, removing correlated variables is likely to be advantageous.

In comparing the two data reduction techniques tested here, reducing original variables but continuing to use original data values (the Reduced-Original dataset), generally performed better than reducing data using PCA (the Reduced-Synthetic dataset). Differences, however, were fairly small, particularly for the best scoring method (Percentile > 0), and generally not different at the .10 level of significance (i.e. χ^2 differences were <3.2). The creation of the Reduced-Original dataset was based primarily on the χ^2 values derived from the independent USEPA classifications. Since most studies that rely on watershed disturbance indices do not have the benefit of independently derived classifications, eliminating redundant variables using χ^2 tests would not be viable. However, given that data reduction using PCA produced indices that performed nearly as well as that from χ^2 analysis, PCA is likely to be a viable alternative in creating a reduced dataset in most studies.

4.2. Effect of scoring method

The Percentile > 0 scoring method outperformed both of the other scoring methods (Fig. 5b). This method differs from the others in two regards: it allows all data values to contribute to the overall index score (Percentile > 50 does not), and it bins data prior to assigning raw disturbance points (Range-Standardize assigns points based on a rescaled 0–10 scheme, allowing for potentially more skew in the results due to outliers). It seems therefore advantageous for an index to incorporate these characteristics in some way. The distribution of index scores from the Percentile > 0 method also provided a more uniform gradation of values (Fig. 6) compared to the other scoring methods. Because the Percentile > 50 method leaves many data

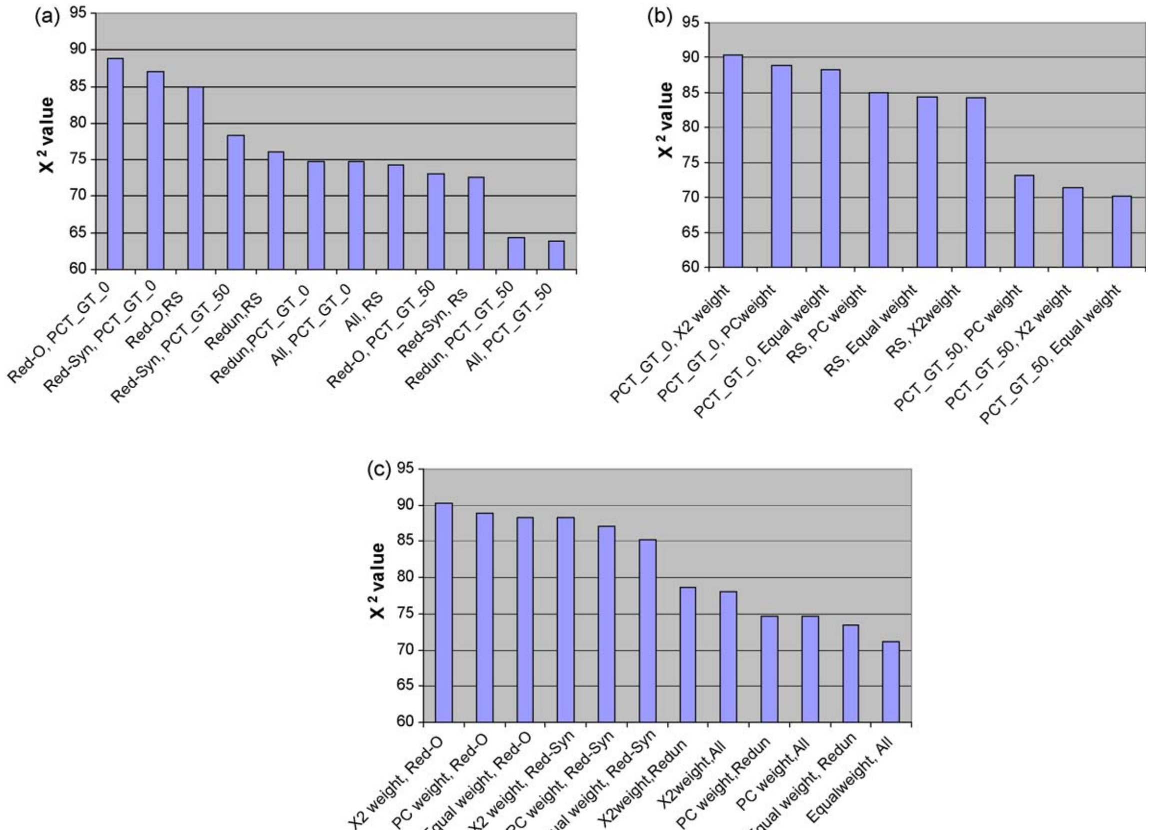

Fig. 5. Comparison of χ^2 values for indices, varying two elements of index creation and holding one element constant. (a) The effect of redundant variables, varying scoring method and holding weighting constant (PC weighting). (b) The effect of scoring method, varying weighting, holding dataset constant (Reduced-Original dataset). (c) The effect of variable weighting, varying dataset and holding scoring method constant (Percentile $>$ 0). Abbreviations for scoring methods: RS = range-standardize, PCT_GT_0 = Percentile $>$ 0 and PCT_GT_50 = Percentile $>$ 50. Abbreviations for datasets: Red-O = Reduced-Original, Red-Syn = Reduced-Synthetic, Redun = redundant and All = all original variables.

values unscored, considerably more sites result in identical 0.0 scores, making them indistinguishable. The Range-Standardize method likewise forced many scores to low values because of the presence of a small number of sites with very high data values. The Percentile $>$ 0 method would be the method of choice if one of the goals of the index was to allow differentiation among sites.

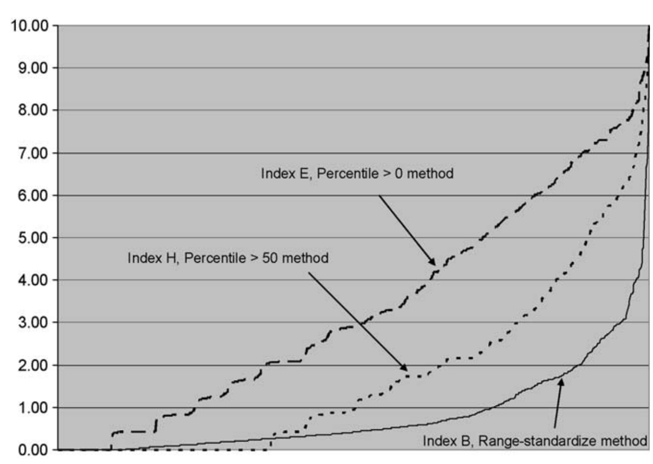

Fig. 6. Comparison of distribution of index scores (y-axis) for three indices. Each index was created with the same input variables in an identical manner (Reduced-Original dataset, χ^2 weighting) except for the scoring method. X-axis represents sites ordered from lowest (0) to highest (10) score for each index (calibration watersheds; n = 770).

4.3. Effect of variable weighting

Both the χ^2 and PC weightings improved indices considerably over equal weighting when used with all 33 variables (i.e. statistically different at at least the 0.10 level), to a lesser degree with the Redundant dataset, and less still with indices based on the two reduced datasets (Fig. 5c). The fact that weighting usually improved the indices only slightly with the two reduced datasets would indicate that the information present in the weightings was already incorporated by reducing the data. We would conclude that weighting the variables is important to the degree that redundancy is present in the final dataset; i.e. it is less important, or even undesirable (if done poorly), if variables have already been well reduced.

What weighting scheme may be applied? In this study the PC-based weighting generally did slightly better than the χ^2 weighting (8 of 12 comparisons). Given that the χ^2 weighting was based on independent information that would ordinarily not be available (how well the variables individually separated USEPA-classified REF and DIS sites), and the PC weighting based entirely on the data themselves, the PC weighting would seem to be a reasonable method for future applications. It may be that other methods of deriving weights from the data themselves (e.g. importance of variables based on clustering), or using best professional judgment may provide good alternate ways of identifying variable importance. It is noteworthy that our *a priori* best professional judgment about which variables in the Reduced-Original dataset should have the highest weights (measures of urbanization and agriculture, over measures of impoundment and flow modification) corresponded generally to the weights they ultimately got in both weighting schemes. This result would suggest that knowledge of the sites themselves and professional judgment is likely to provide a reasonably good alternative to data-driven weighting.

4.4. Multimetric index versus individual variables

Although some of our variables were composites (e.g. URBCP_MAINS integrated urban and agriculture land cover), no single variable (values in Table 3) was nearly as successful as the vast majority of multimetric indices tested in this study (Table 5). The worst of the Reduced-Original indices tested (index G; $\chi^2 = 70.2$) was better than the best individual variable tested (HUDEN; $\chi^2 = 66.9$), as judged by ability to distinguish REF and DIS sites.

Although no single variable performed as well as the majority of our indices, a few of our multimetric indices performed worse than individual variables. For example, a variation using all 33 original variables (index AK in Table 5; $\chi^2 = 54.2$) was outperformed by a number of individual measures of urbanization and agriculture. However, by incorporating some reasonably simple rules in index creation – removing redundant variables, scoring all data values, employing some logical weighting scheme – a disturbance index is likely to be a considerably more powerful way of quantifying anthropogenic disturbance than using any single landscape factor.

4.5. Regionalization

One aspect of index creation that became evident as the indices were tested was the difference between index scores by region (Fig. 4). By any index method tested, sites in the Plains region clearly had the highest amount of anthropogenic disturbance, followed by the Xeric then Mountains regions. Our method of assigning scores did not take into account the site's region: disturbance points were assigned purely on variables' data values. However, when selecting thresholds for classification, it is clearly important that sites be compared to sites from the same region. As has been noted elsewhere (Whittier et al., 2007) and was seen here, least-disturbed sites from a highly urbanized or mechanized region (e.g. Plains) may have higher disturbance scores than most-disturbed sites from a more pristine region (e.g. Mountains).

4.6. Watershed classification

There are a number of obstacles to correctly categorizing or ranking watersheds over a broad area using GIS-derived data. First, the GIS data themselves are derived from different sources and agencies, are in different formats (e.g. continuous surface land cover data versus vector line data representing roads versus point locations representing dams), may represent varying time periods which are older than the dependent variable data, and in every case contain inaccuracy. In some cases the accuracy is roughly known (e.g. land cover; Stehman et al., 2003), but in most cases accuracy is unknown (for example, dam or point-source location data). If accuracy was known and comparable for all data layers the index calculation could be adjusted accordingly (for example, down-weighting less accurate datasets), however, the reality is that a certain amount of unknown noise is added to the index based on the underlying data. Given that even moderate amounts of error may hinder identification of patterns in geographic data (Jacquez and Waller, 1997), the issue of uncertainty in the GIS data is likely to be significant in identifying subtle differences among watersheds.

Second, even if accurate, nationally available GIS coverages may lack the resolution required to discern the presence of human activities. This was illustrated in our study, where watersheds classified by USEPA using a combination of site measurements (e.g. water chemistry and stream habitat) and aerial imagery apparently lacked evidence of human disturbance in the GIS data. For example, 42% of the watersheds classified by USEPA as most-disturbed had <1% agriculture and <1% urban land cover, and 2/3 of those records had complete absence of dams, canals/ditches/pipelines, pollutant discharge locations, mining land cover, and mineral operations locations. In short, 28% of the watersheds that USEPA classified as most-disturbed lacked any disturbance characteristic in our GIS data. There is no doubt that fine-scale GIS data for such disturbances as localized dredging, channelization, presence of livestock, construction, or logging would have benefited the indices we created. These were not available to us at the scale of this study, and remain possibly the greatest drawback to national or regional implementation of an anthropogenic disturbance index based on relatively coarse-scale GIS data.

Despite the weaknesses of a disturbance index derived from national-scale GIS data, our results suggest there may be important advantages to such an index. Although limited by the coarse resolution of national-scale GIS coverages, our index correctly classified more than 2/3 of an independent set of sites as either least- or most-disturbed. For regional and national assessments where the cost of obtaining high resolution imagery and site data are prohibitively high, a disturbance index based solely on GIS data may be highly beneficial given that the data are already available. In addition, a nationally consistent index of watershed disturbance could be used to screen a large number of watersheds to identify streams that are likely to be in relatively undisturbed condition or, conversely, streams that are potentially exposed to excessive anthropogenic disturbance.

5. Conclusion

In this study we derived disturbance index scores that were used to rank Western watersheds and we attempted to identify strategies for separating least- from most-disturbed sites. We evaluated how decisions in index creation (e.g. data reduction, variable scoring, and weighting) influenced overall performance. While the variable-specific results given here (which variables were most discriminating) may be specific to the western United States, we believe that the aspects of index creation in general are applicable to any region. We suggest that these procedures will likely improve performance of a disturbance index: (1) removing redundant variables, (2) scoring all data values, and (3) incorporating a logical weighting scheme. Of these three aspects of index creation, the largest improvements in index performance came from data reduction, followed by scoring method, and the least improvement from implementation of weighting.

In this study data reduction to synthetic variables based on multivariate statistical analysis (PCA) provided nearly as good a result as selection of original variables based on external information. Of the three scoring techniques we tested, a method which scored all data values by binning them to percentile-based categories proved the most effective. The multimetric indices developed for this study were overall much more effective in separating least- and most-disturbed sites than any individual variable, even though a number of our variables were composites which combined several disturbance factors (e.g. urban and agricultural land).

The main conclusion of this study is not in the details of how an index may be created, or that certain methods are necessarily better, but that (a) a disturbance index is likely to be a better screening tool for identifying least- and most-disturbed sites than an individual variable or class of variables (e.g. land cover), and that (b) the method of index creation can make a difference. It is recognized that there are difficulties in implementing a GIS-based disturbance index for small-to-medium sized watersheds over broad regions because of the difficulty in obtaining consistent and detailed information. It is clear that the success of the disturbance index will be dependent on the accuracy and resolution of the GIS data themselves. Nonetheless, as was shown here, thoughtful implementation of the method for creating an index can make a difference in results, even at a broad, regional scale.

Acknowledgements

Grateful acknowledgment to Ryan Hill and Chuck Hawkins of Utah State University for help in this project and providing initial data. Many thanks also to Gary Buell (USGS) and Alan Herlihy (USEPA) for very helpful review comments and suggestions.

References

Allan, J.D., 2004. Landscapes and riverscapes: the influence of land use on stream ecosystems. Annual Review of Ecology, Evolution, and Systematics 35, 257–284.

Blocksum, K.A., 2003. A performance comparison of metric scoring methods for a multimetric index for mid-Atlantic highlands streams. Environmental Management 31, 670–682.

Brown, M.T., Vivas, B., 2005. Landscape development intensity index. Environmental Monitoring and Assessment 101, 289–309.

Danz, N.P., Regal, R.R., Niemi, G.J., Brady, V.J., Hollenhorst, T., Johnson, L.B., Host, G.E., Hanowski, J.M., Johnston, C.A., Brown, T., Kingston, J., Kelly, J.R., 2005. Environmentally stratified sampling design for the development of great lakes environmental variables. Environmental Monitoring and Assessment 102, 41–65.

Danz, N.P., Niemi, G.J., Regal, R.R., Hollenhorst, T., Johnson, L.B., Hanowski, J.M., Axler, R.P., Ciborowski, J.J.H., Hrabik, T., Brady, V.J., Kelly, J.R., Morrice, J.A., Brazner, J.C., Howe, R.W., Johnston, C.A., Host, G.E., 2007. Integrated Measures of Anthropogenic Stress in the U.S. Great Lakes Watershed. Environmental Monitoring and Assessment 39, 631–647.

Falcone, J.A., Stewart, J.S., Sobieszczyk, S., Dupree, J.A., McMahon, G., Buell, G.R., 2007. A comparison of natural and urban characteristics and the development of urban intensity indices across six geographic settings. U.S. Geological Survey Scientific Investigations Report 2007-5123.

GeoLytics, 2001. Census 2000 and Street 2000. GeoLytics, Inc., East Brunswick, NJ (2 CDROMS).

Host, G.E., Schuldt, J., Ciborowski, J.J.H., Johnson, L.B., Hollenhorst, T., Richards, C., 2005. Use of GIS and remotely sensed data for a priori identification of reference areas for Great Lakes coastal ecosystems. International Journal of Remote Sensing 26, 5325–5342.

Jacquez, G.M., Waller, L.A., 1997. The effect of uncertain locations on disease cluster statistics. In: Mowerer, H.T., Congalton, R.G. (Eds.), Quantifying Spatial Uncertainty in Natural Resources: Theory and Application for GIS and Remote Sensing. Arbor Press, Chelsea, MI, USA.

Karr, J.R., Chu, E.W., 1999. Restoring Life in Running Waters: Better Biological Monitoring. Island Press, Washington, DC.

Knepp, D.L., Entwisle, D.R., 1969. Testing significance of differences between two chi-squares. Psychometrika 34, 331–333.

McMahon, G., Cuffney, T.F., 2000. Quantifying urban intensity in drainage watersheds for assessing stream ecological conditions. Journal of the American Water Resources Association 36, 1247–1261.

Price, C.P., Nakagaki, N., Hitt, K.J., Clawges, R.M., 2006. Enhanced historical land-use and land-cover datasets of the U.S. Geological Survey. U.S. Geological Survey Data Series 240, digital maps, accessed in July 2007 at http://pubs.usgs.gov/ds/2006/240.

Ruddy, B.C., Lorenz, D.L., Mueller, D.K., 2006. County-level estimates of nutrient inputs to the land surface of the conterminous United States, 1982–2001. U.S. Geological Survey Scientific Investigations Report 2006-5012.

Sanderson, E.W., Jaiteh, M., Levy, M.A., Redford, K.H., Wannebo, A.V., Woolmer, G., 2002. The human footprint and the last of the wild. Bioscience 52, 891–904.

Stehman, S.V., Wickham, J.D., Smith, J.H., Yang, L., 2003. Thematic accuracy of the 1992 National land-cover data (NLCD) for the eastern United States: statistical methodology and regional results. Remote Sensing of Environment 86, 500–516.

Stein, J.L., Stein, J.A., Nix, H.A., 2002. Spatial analysis of anthropogenic river disturbances at regional and continental scales. Landscape and Urban Planning 60, 1–25.

U.S. Army Corps of Engineers, 2006. National Inventory of Dams. U.S. Army Corps of Engineers Accessed in July 2006 at http://crunch.tec.army.mil/nidpublic/webpages/nid.cfm.

U.S. Environmental Protection Agency, 2005. Western Streams and Rivers Statistical Summary. U.S. Environmental Protection Agency EPA 620/R-05/006. Accessed in October 2007 at http://crunch.tec.army.mil/nidpublic/webpages/nid.cfm.

U.S. Environmental Protection Agency, 2006a. Wadeable Streams Assessment: A Collaborative Survey of the Nation's Streams. U.S. Environmental Protection Agency EPA 841-B-06-002 December 2006. Accessed in October 2007 at http://www.epa.gov/owow/streamsurvey/

U.S. Environmental Protection Agency, 2006b. National Pollutant Discharge Elimination System (NPDES)U.S. Environmental Protection Agency Accessed in June 2006 at http://cfpub.epa.gov/npdes/

U.S. Environmental Protection Agency, 2006c. National Hydrography Dataset Plus (NHDPlus) Home PageUSEPA, USGS, and Horizon Systems Corporation Accessed in August 2006 at http://www.horizon-systems.com/nhdplus/

U.S. Geological Survey, 2007. Grids of Agricultural Pesticide Use in the Conterminous United States 1997. U.S. Geological Survey Accessed in June 2007 at http://water.usgs.gov/GIS/metadata/usgswrd/XML/agpest97grd.xml.

U.S. Geological Survey, 2006. National Atlas Home page. U.S. Geological Survey Accessed in November 2006 at http://www-atlas.usgs.gov.

U.S. Geological Survey, 2002. Open File Report 041189607. U.S. Geological Survey Accessed in August 2007 at http://water.usgs.gov/GIS/metadata/usgswrd/XML/ofr041189607.xml.

Walpole, R.E., Myers, R.H., 1978. Probability and Statistics for Engineers and Scientists. Macmillan Publishing, New York, NY, USA.

Wang, L., Brenden, T., Seelbach, P., Cooper, A., Allan, D., Clark Jr., R., Wiley, M., 2008. Landscape based identification of human disturbance gradients and reference conditions for Michigan streams. Environmental Monitoring and Assessment 141, 1–17.

Whittier, T.R., Stoddard, J.L., Larsen, D.P., Herlihy, A.T., 2007. Selecting reference sites for stream biological assessments: best professional judgment or objective criteria. Journal of the North American Benthological Society 26, 349–360.

Wilhelm, J.G.O., Cummins, K.W., Allan, J.D., Wessell, K.J., Merritt, R.W., 2005. Habitat assessment of non-wadeable rivers in Michigan. Environmental Monitoring and Assessment 36, 592–609.

Yang, K., Shahabi, C., 2004. A PCA-based similarity measure for multivariate time series. In: MMDB'04: Proceedings of the 2nd ACM International Workshop on Multimedia Databases. ACM Press, Washington, DC, USA, pp. 65–74, Accessed in December 2007 at http://infolab.usc.edu/DocsDemos/mmdb04.pdf.